Taming Big Data to Create Usable Knowledge

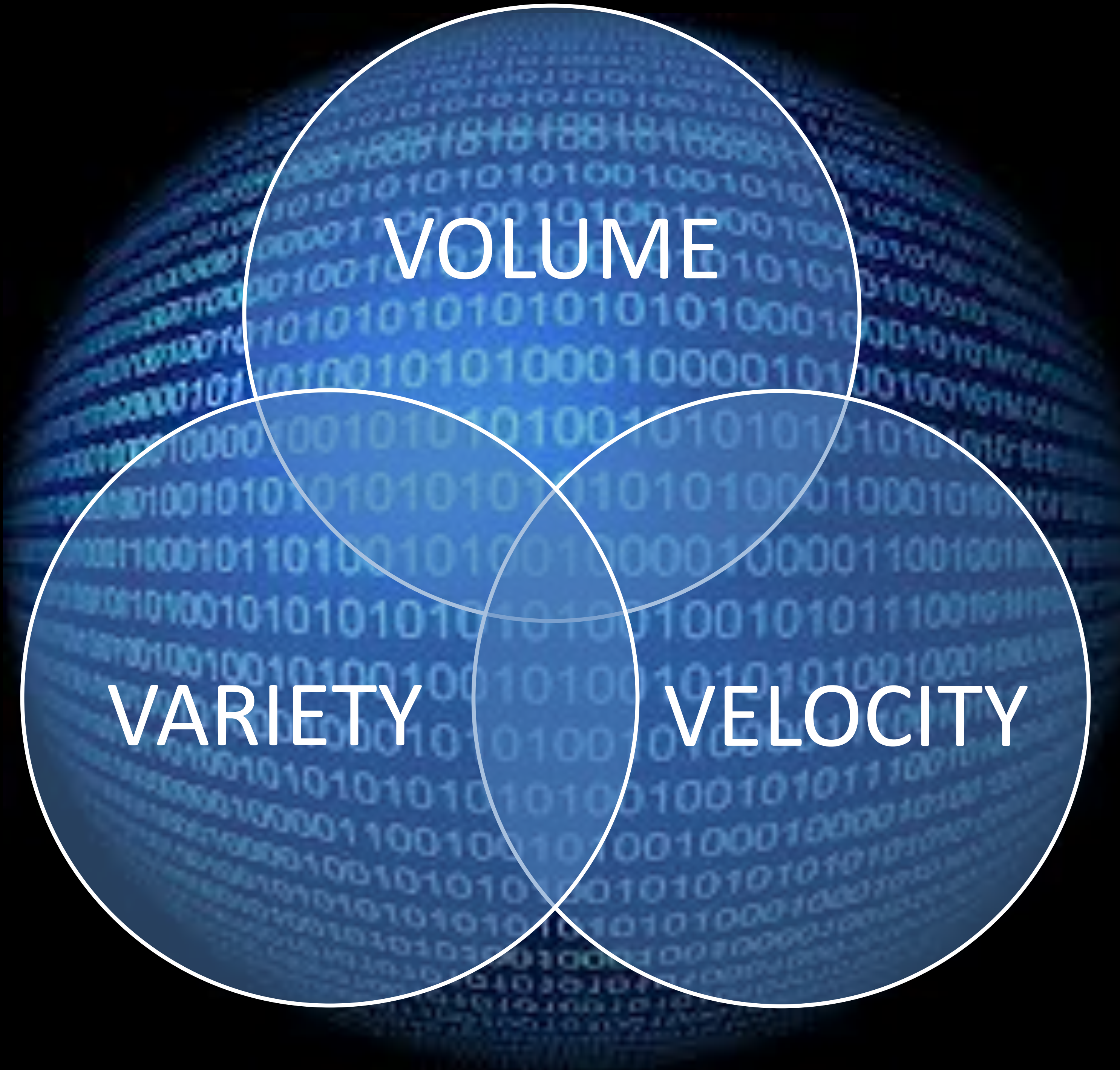
Denise Lach
Oregon State University

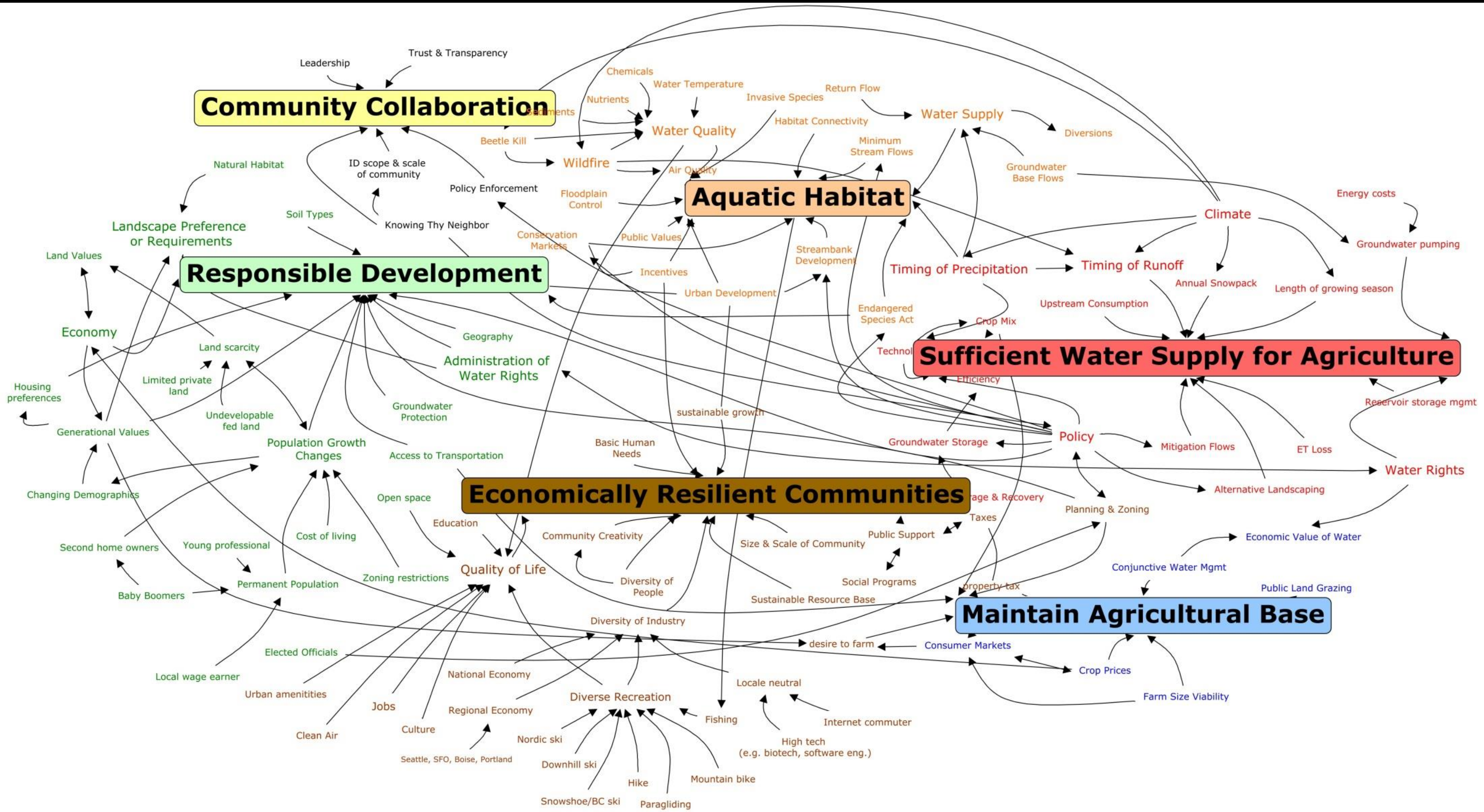
Big data is the broad name given to challenges and opportunities we have as data about every aspect of our lives becomes available. It’s **not just about data though; it also includes the people, processes, and analysis that turn data into meaning.**

John Ferguson, Chief Technology Officer
Mode Analytics

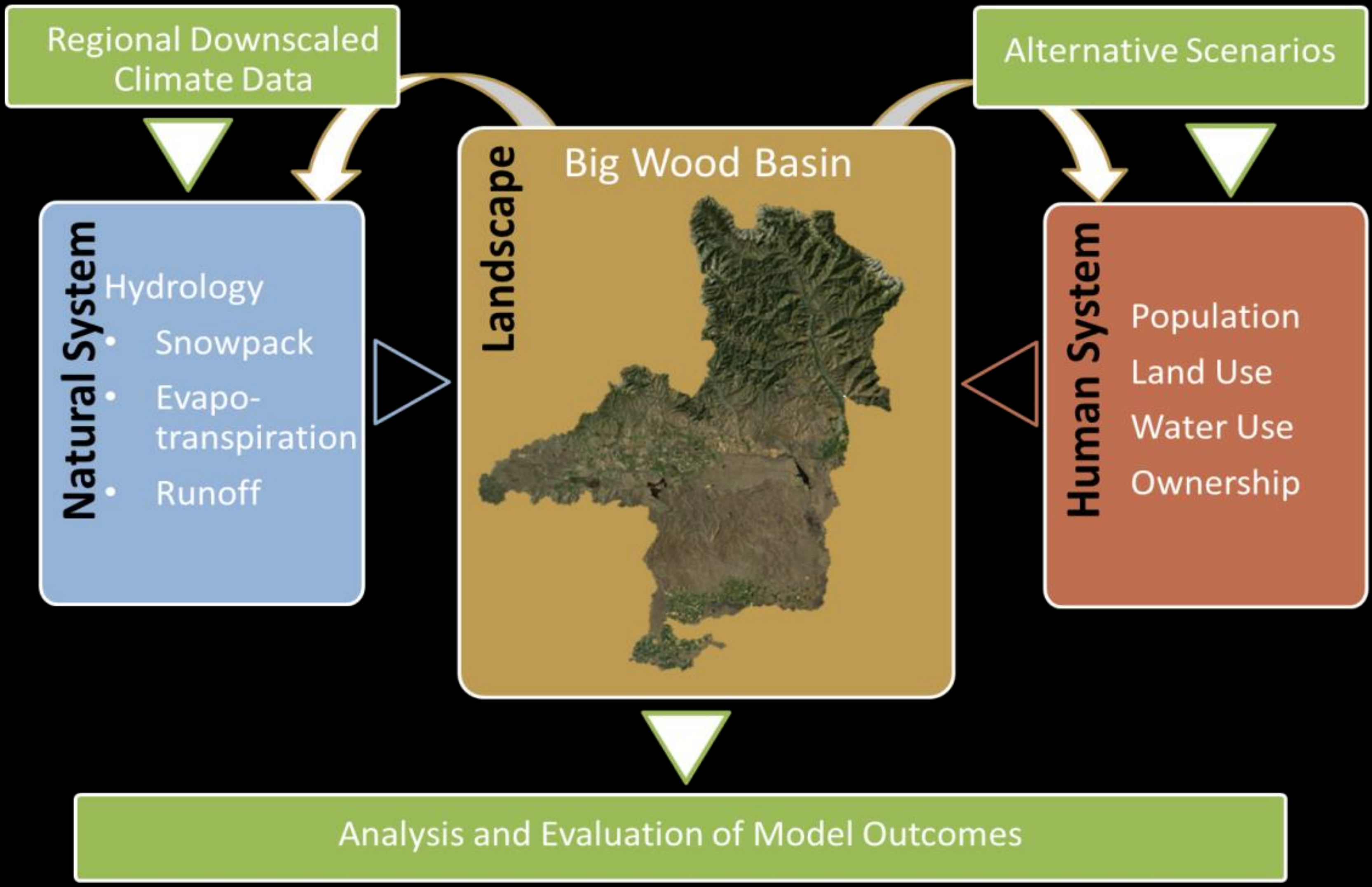
For me, the technological definitions (like "too big to fit in an Excel spreadsheet" or "too big to hold in memory") are important, but aren't really the main point. Big data for me is data at a scale and scope that changes in some fundamental way (not just at the margins) the range of solutions that can be considered when people and organizations face a complex problem. Different solutions, not just 'more, better.'

Steven Weber, Professor
UC Berkeley, School of Information
Department of Political Science

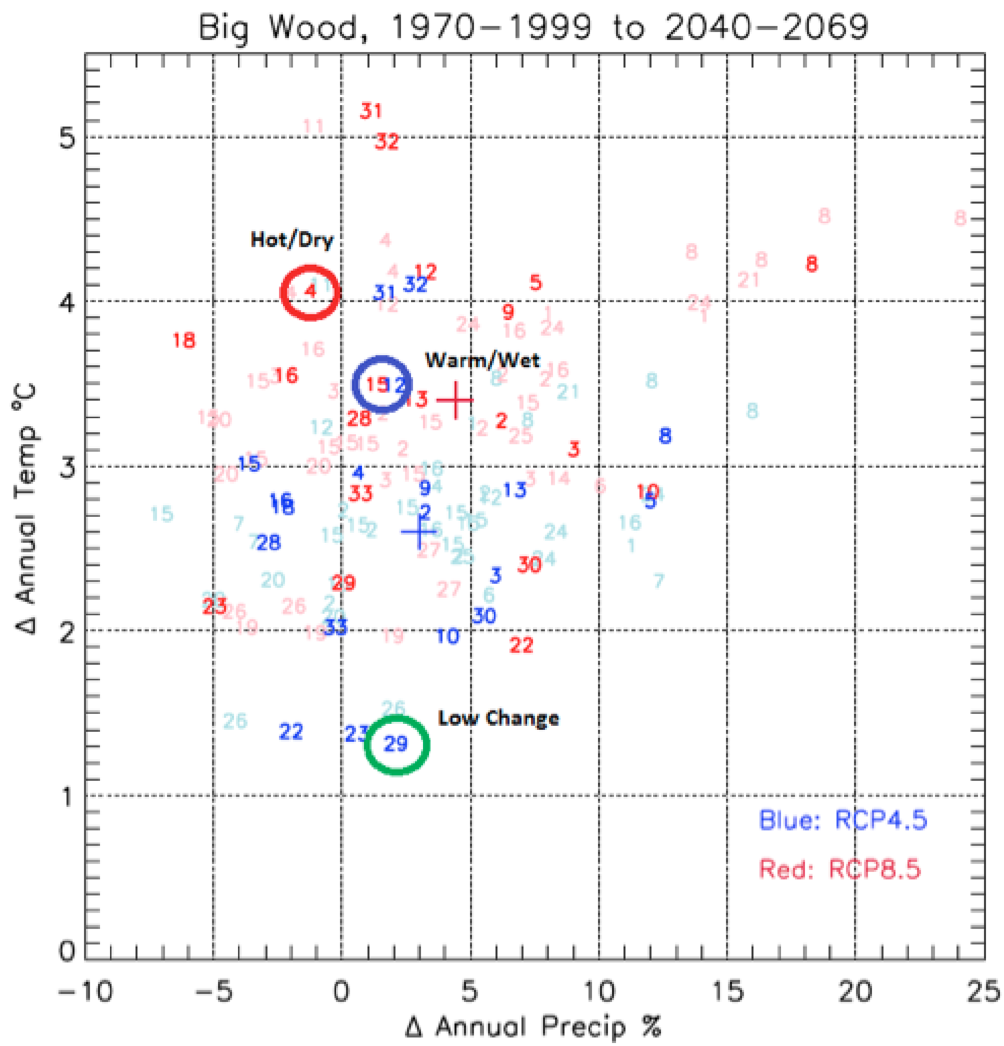

VOLUME
VARIETY
VELOCITY


Community Collaboration
Leadership
Trust & Transparency
ID scope & scale of community
Policy Enforcement
Knowing Thy Neighbor
Natural Habitat
Soil Types
Landscape Preference or Requirements
Land Values
Responsible Development
Economy
Land scarcity
Geography
Administration of Water Rights
Housing preferences
Limited private land
Undevelopable fed land
Groundwater Protection
Generational Values
Population Growth Changes
Access to Transportation
Changing Demographics
Open space
Second home owners
Young professional
Cost of living
Education
Baby Boomers
Permanent Population
Zoning restrictions
Quality of Life
Elected Officials
Local wage earner
Urban amenities
Clean Air
Jobs
Culture
Seattle, SFO, Boise, Portland
National Economy
Regional Economy
Chemicals
Water Temperature
Nutrients
Sediments
Water Quality
Invasive Species
Return Flow
Water Supply
Diversions
Beetle Kill
Wildfire
Air Quality
Habitat Connectivity
Minimum Stream Flows
Groundwater Base Flows
Floodplain Control
Aquatic Habitat
Conservation Markets
Public Values
Incentives
Urban Development
Streambank Development
Endangered Species Act
Climate
Energy costs
Groundwater pumping
Timing of Precipitation
Timing of Runoff
Annual Snowpack
Length of growing season
Upstream Consumption
Crop Mix
Technol
Sufficient Water Supply for Agriculture
Efficiency
Reservoir storage mgmt
sustainable growth
Basic Human Needs
Groundwater Storage
Policy
Mitigation Flows
ET Loss
Water Rights
Economically Resilient Communities
rage & Recovery
Alternative Landscaping
Planning & Zoning
Taxes
Community Creativity
Size & Scale of Community
Public Support
Economic Value of Water
Conjunctive Water Mgmt
Public Land Grazing
Diversity of People
Social Programs
Sustainable Resource Base
property tax
Maintain Agricultural Base
Diversity of Industry
desire to farm
Consumer Markets
Crop Prices
Farm Size Viability
Diverse Recreation
Locale neutral
Fishing
Internet commuter
Nordic ski
Downhill ski
High tech (e.g. biotech, software eng.)
Hike
Mountain bike
Snowshoe/BC ski
Paragliding

ENVISION Model Framework



Big Wood Climate Model Selection



1. CESM1–CAM5
2. CCSM4
3. CNRM–CM5
4. HadGEM2–ES
5. HadGEM2–CC
6. CMCC–CM
7. EC–EARTH
8. CanESM2
9. IPSL–CM5A–MR
10. bcc–csm1–1–m
11. HadGEM2–AO
12. MIROC5
13. NorESM1–M
14. CMCC–CMS
15. CSIRO–Mk3–6–0
16. IPSL–CM5A–LR
17. MPI–ESM–MR
18. BNU–ESM
19. FIO–ESM
20. MPI–ESM–LR
21. GFDL–CM3
22. MRI–CGCM3
23. inmcm4
24. FGOALS–s2
25. FGOALS–g2
26. GISS–E2–R
27. GISS–E2–H
28. bcc–csm1–1
29. GFDL–ESM2M
30. GFDL–ESM2G
31. MIROC–ESM–CHEM
32. MIROC–ESM
33. IPSL–CM5B–LR

Δ Temperature, °C

	RCP4.5	RCP8.5
Max.	4.1	5.1
75%	2.9	4.1
Mean	2.6	3.4
25%	2.1	2.8
Min.	1.3	1.9

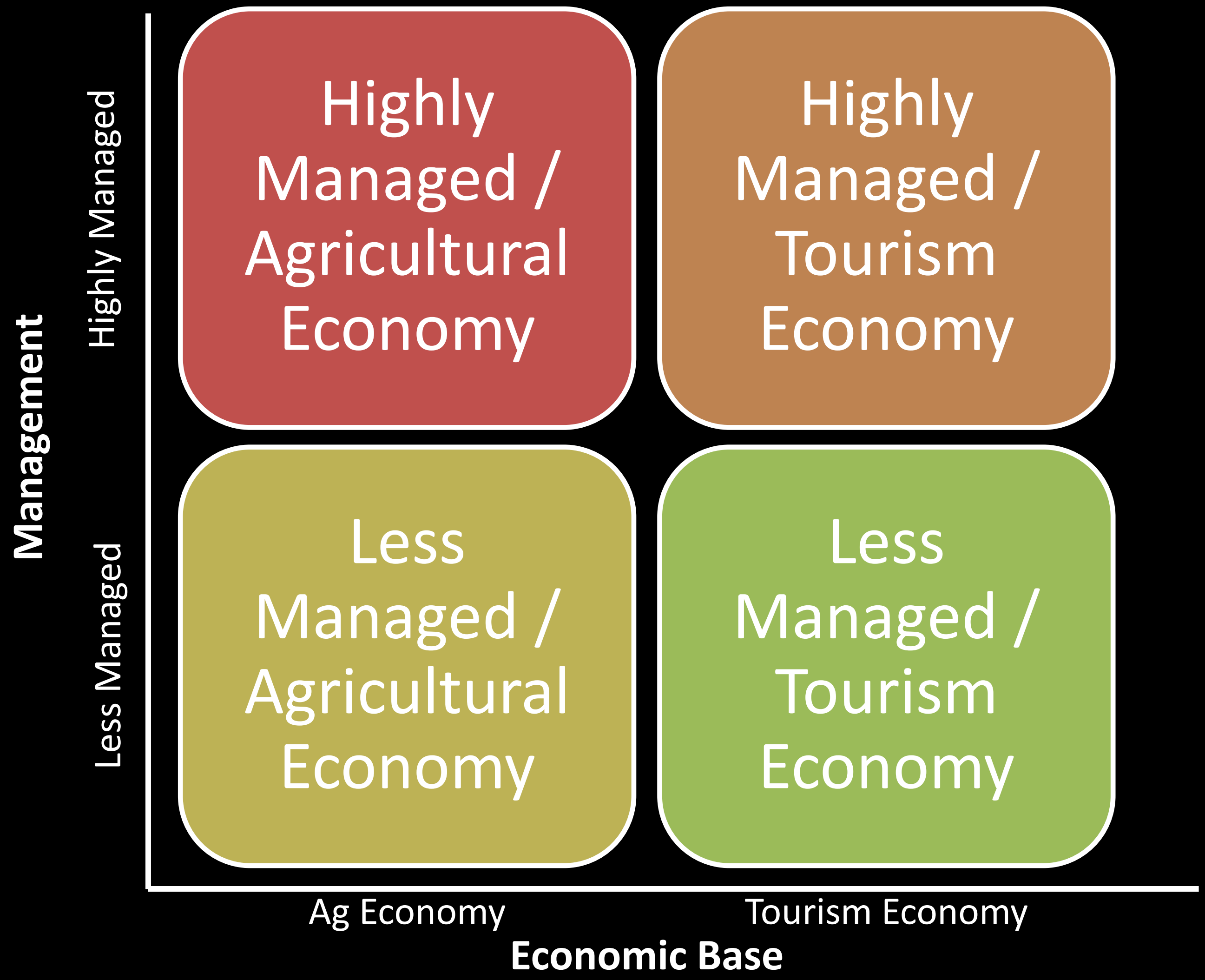
Δ Precipitation, %

	RCP4.5	RCP8.5
Max.	12.1	18.3
75%	5.8	7.1
Mean	3.0	4.4
25%	−0.20	0.20
Min.	−2.8	−6.1

Where models had multiple ensemble members, ensembles were averaged prior to calculating statistics in the table above.

Key:
Dark color = MACA available
Light color = MACA not available
Models ranked by performance index
"+" = mean of RCP

Alternative Scenarios: Economic base, management approach



12 Alternative Scenarios: economic base, management approach, climate scenario

Communicating Complicated Information: What's Important?

Envision Big Wood Basin

Exploring water futures under alternative climate and management scenarios

Home | Introduction | Scenarios | Storylines | Conclusions | About

Background - The Big Wood Basin

<<Previous: Table of Contents   Next: The Process>>

The site describes an "alternative futures" assessment for the Big Wood Basin, Idaho.

Location

The total study area encompasses the Big Wood River, Little Wood River, and Camas Creek drainages in central Idaho totaling approximately 8,300 square kilometers (see Figure 1). The primary focus of the study is on the water resources of the Big Wood River basin, which includes the Big Wood River and Camas Creek (approx. 6,000 square kilometers). However, approximately 100 square kilometers in the Little Wood River drainage are irrigated from the Big Wood River so that basin is included in the study area for the purpose of simulating irrigation although otherwise it is not studied in detail. The study area lies within portions of Blaine, Camas, Elmore, Gooding and Lincoln counties and the major population centers include Ketchum, Sun Valley, Hailey, Bellevue, Fairfield, and Gooding.

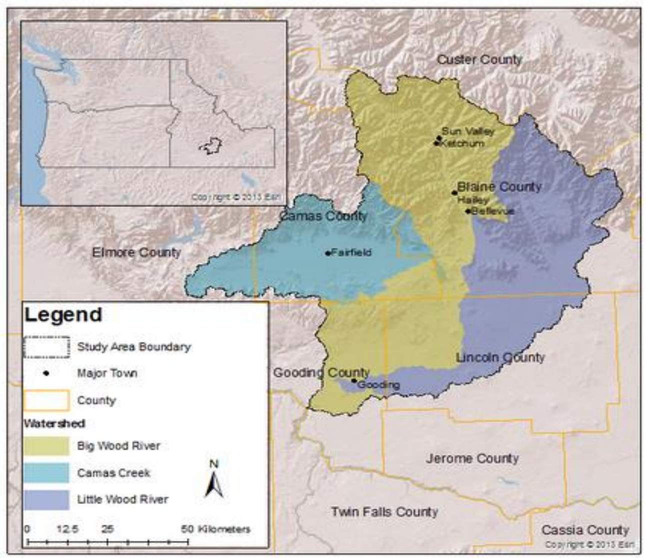


Figure 1. Map of Study Area.

Land ownership in the study area is approximately 66% public and 34% private. The majority (58%) of the public land is managed by the U.S. Bureau of Land Management (BLM) with the U.S. Forest Service overseeing 35% and the State of Idaho managing approximately 5% (U.S. Bureau of Land Management, Idaho State Office, Geographic Sciences, 2009).

[Big Wood Data Atlas]

Communicating Usable Knowledge

- Transparent assumptions
- Important issues/questions
- Simple visuals
- Multiple options for individual exploration
- Intuitive interface
- Meta data and data access

Questions?

Lessons Learned

- Need lots of time
- Local expertise and data increase model legitimacy
- Lots to learn about understanding and communicating complicated model results
- Conversations about future not caught in arguments about what/when/if/who

Storyline - Stream Flows

<<Previous: Magic Reservoir   Next: Land Use>>

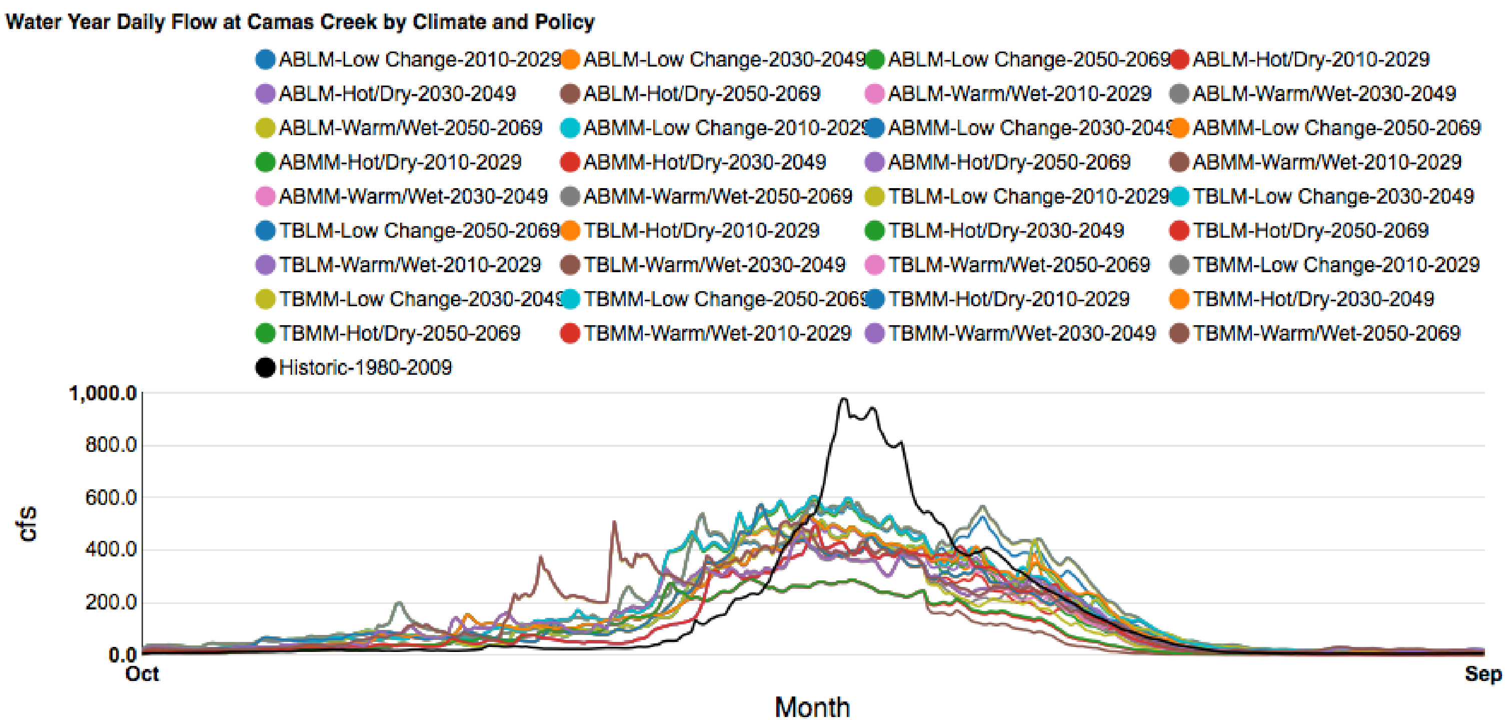
Take Home Message

Background

The **timing** and **magnitude** of stream flow in the Big Wood basin is controlled by a number of factors - climate, snowpack, timing of releases from Magic Reservoir, surface water/groundwater interactions, demand from natural vegetation, agricultural and urban uses, among others. Most of these factors (except surface water/groundwater interactions) are consider in the scenario models.

The results below show some of the results related to stream flows. The first and second chart show the annual stream flow profile for various decades for Camas Creek and the Big Wood at Hailey for each of the twelve management X climate scenarios. There is a lot of data on this chart, so it may be helpful to use the filters to explore these.

Water Year Average Flows (cfs) at Camas Creek



Filter: [ ] reset Shortcuts: ablm abmm tblm tbmm Low Change Warm/Wet Hot/Dry

Envision Big Wood Basin

Exploring water futures under alternative climate and management scenarios

Storyline - Snowpack

<<Previous: Precipitation   Next: SWE Maps>>

Take Home Message

Snow in the Big Wood Basin is an important storage mechanism satisfying late spring and summer water demand lower in the basin. The climate models show clear trends in the **timing** of snow, with more of a mixed story regarding the **amount** of snow expected into the future. In general, the amount of snow tends to continue the current downward trend, but there is some decade variation in this. These impacts of reduced (and earlier) snow are addressed later on in this storyline.

Key Findings

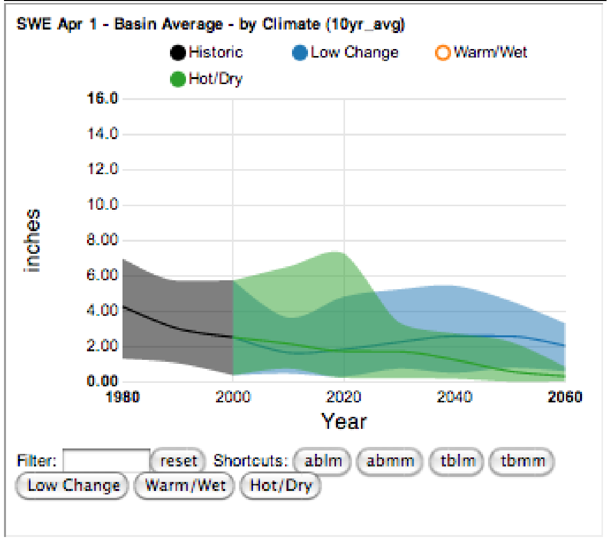
- Snowpack has declined over the last 30 years.
- It is difficult to identify a consistent future trend in the volume of peak snowpack across the scenarios.
- However, all scenarios suggest a change in the timing of the peak seasonal snow (historically found near April 1) which may occur up to 6 weeks earlier.
- High elevation snowpack, where most of the basin's snow occurs, will be most impacted.

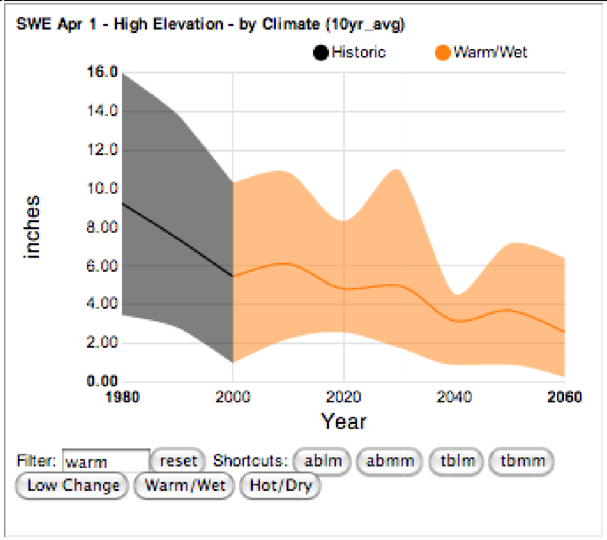
Background

Snow is critical to the area, because it represents water **storage** in the system, similar to a reservoir and supports winter recreation. Loss of snowpack implies loss of that storage capacity, which is either gone or requires replacement in the form of man-made reservoirs. We show both the amount of snow, and the timing of the snow, to capture this storyline. Snowpack is expressed here as **"Snow Water Equivalents"**, or **SWE** - the amount of water contained in snow - and is sensitive to changes in both temperature and timing of precipitation.

The Data

The following chart shows a summary of the basin average snow (SWE) on April 1 for the entire Big Wood Basin.	The chart below shows similar data, but only for the high elevation (> 6500ft) areas in the basin, for each of the climate scenarios.





Storyline - Water Demand

<<Previous: Snowpack   Next: Agriculture>>

Take Home Message

Adoption of policies that address water efficiency make a difference. This is particularly true regarding agricultural water use.

Key Findings

- Water demand varies widely by sector and by management scenario. Because of it's extensive area in the basin, **natural shrubland** is estimated to be the largest "consumer" of water in the basin, followed by **agriculture**.
- It is also clear from these results that for "agriculture boom" scenarios, **management matters**; there are quite marked differences between less managed and more managed scenarios. In the less managed scenarios, agricultural water demand is projected to increase by roughly 50%, while under the more managed scenario, agricultural water demand stays roughly constant, despite increasing temperatures and a robust agricultural sector. In the "tourist boom" scenarios, agricultural water use either stays flat (less managed) or decreases (more managed).

Background

Water demand in the basin is reported for each section - Agriculture, Forests, Natural Shrubland, and Municipal Uses.

- **Agricultural** water demand focuses on the water used to support crop production, including irrigation. Agricultural water use is determined based on evapotranspiration of the crops, which is determined based on crop type, weather, soil type, and soil moisture status. A description of the methodology used to estimate crop water demand here
- **Forest** and **natural shrubland** water demand is based on estimates of evapotranspiration, estimated in a similar manner to that of crops.

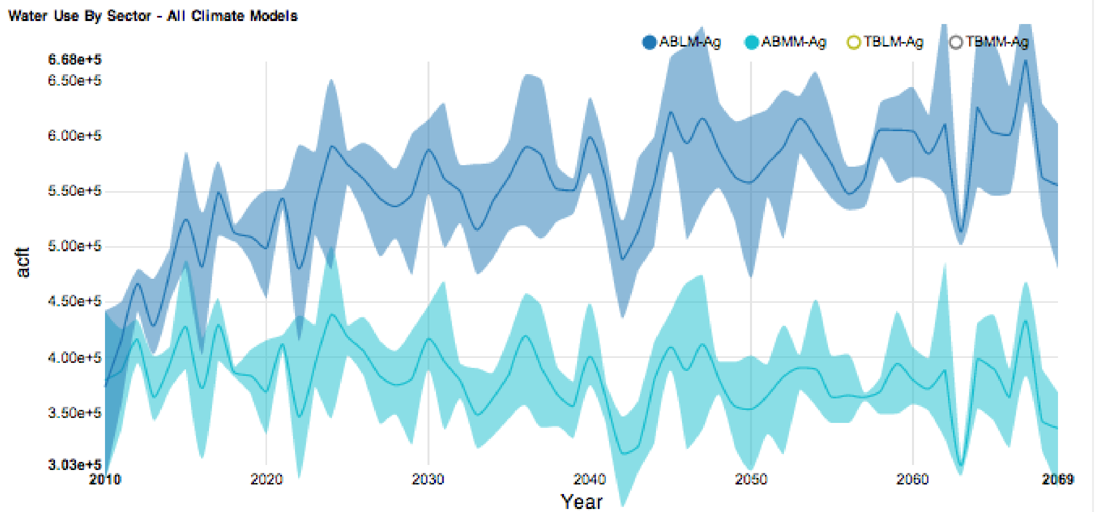
The Data

Results, expressed as percent of total water demand, are presented in the table to right, for current conditions (ca. 2010) and with ranges under the different scenarios for 2070.

Land Cover Type	Percent of Total Demand	
	2010	2070
Natural Shrub	49	47-54
Agriculture	26	20-31
Forest	25	22-26

The following chart shows estimated water demand by sector for the various scenarios. It is helpful to "filter" the information presented on the chart to get a better understand of the effect of different management scenarios of the distribution of water demand.

Annual Water Demand by Sector 2010-2070



Filter: ag [reset] Shortcuts: [ablm] [abmm] [tblm] [tbmm] [Low Change] [Warm/Wet] [Hot/Dry]

This work is supported by NOAA's Climate Program Office with funding from the Regional Integrated Sciences & Assessments (RISA) program

The project is being conducted by the **Climate Impacts Research Consortium (CIRC)**

<<Previous: Model Summary Next: Climate Model Selection>>

Ag Boom – More Managed

(top)

Water Use	Under an agricultural boom – more managed scenario, 25,000 additional ac-ft of storage in Magic Reservoir is assumed and irrigation conveyance efficiency increases by 10%. Agricultural systems shift to more drought tolerant crops. Conservation practices are required for all municipal and private uses.
Land Use	Increased zoning and easements protect prime farmland. In urban areas, effort will be placed to infill development emphasizing higher density in and around existing towns. Ninety percent of new population growth will be allocated to urban areas. Buffer zones are put in place to separate agricultural land from residential areas.

Ag Boom – Less Managed

(top)

Water Use	Under an agricultural boom - less managed scenario, efficiency of water use in agriculture remains at the status-quo. Agriculture shifts to more water intensive cropping patterns and the amount of water storage remains at the status quo.
Land Use	This scenario assumes significant expansion of agriculture through conversion of non-agricultural private and public land. In rural areas, fewer development constraints allow for more rural residential, commercial, and industrial development. Roughly half of new population growth is allocated to urban areas and half to rural areas.

Tourism Boom – More Managed

(top)

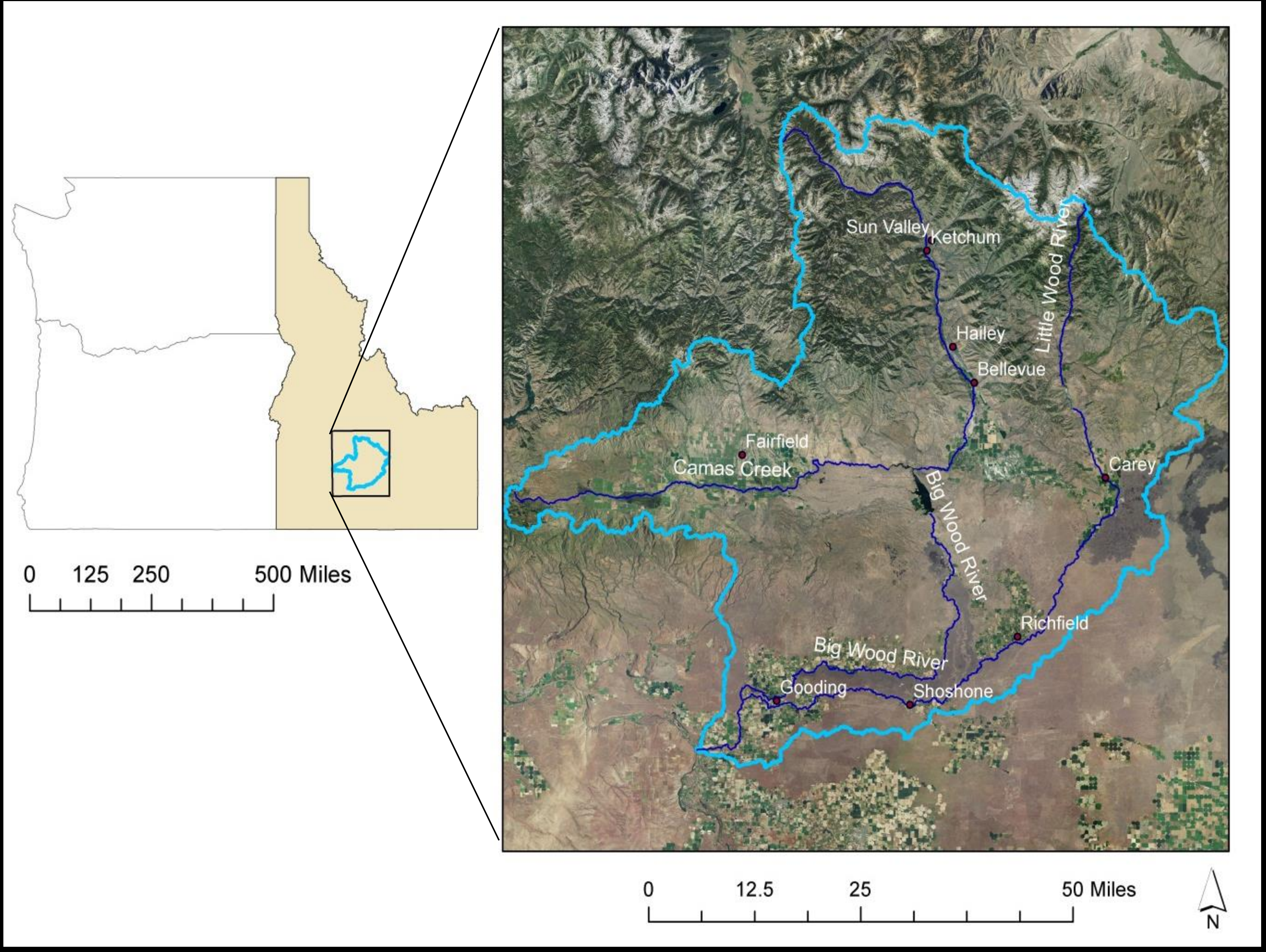
Water Use	Under a tourism boom – more managed scenario, policies to promote high efficiency water use are assumed. Water use in agriculture becomes more efficient. Agricultural systems shift to more drought tolerant crops. 25,000 ac-ft of additional storage in Magic Reservoir is assumed. Conservation practices are required for all municipal and private uses.
Land Use	This scenario assumes limited conversion of agricultural land to conservation reserves, resorts, and spas. Increased zoning and easements protect prime farmland. No new development on public lands is allowed. Ninety percent of new population growth will be allocated to existing urban growth areas. Buffer zones are put in place to separate agricultural land from residential areas. Incentives to convert private lands into conservation uses are assumed such as increased riparian buffers, and wetland setbacks.

Tourism Boom – Less Managed

(top)

Water Use	Under a tourism boom – less managed scenario, residential and agricultural water efficiency policies remain at the status quo. Existing agricultural crop patterns are maintained. No changes to water storage are included.
Land Use	This scenario assumes significant conversion of agricultural land to conservation reserves, resorts, and spas will take place. Limited zoning and easement protections will be provided to prime farmland. Some public lands will be developed as resorts. Urban density will remain at the status quo, while urban boundaries will be expanded. In rural areas, fewer development constraints will allow for more rural residential, commercial, and industrial development.

<<Previous: Model Summary Next: Climate Model Selection>>



0
125
250
500 Miles
Sun Valley
Ketchum
Little Wood River
Hailey
Bellevue
Fairfield
Camas Creek
Carey
Big Wood River
Richfield
Big Wood River
Gooding
Shoshone
0
12.5
25
50 Miles
N

Too much information

Storylines that matter

- Temperature
- Snowpack
- Water Demand

Envision Big Wood Basin

Exploring water futures under alternative climate and management scenarios

Home | Introduction | Scenarios | Storylines | Conclusions | About

Background - The Big Wood Basin

<<Previous: Table of Contents   Next: The Process>>

The site describes an "alternative futures" assessment for the Big Wood Basin, Idaho.

Location

The total study area encompasses the Big Wood River, Little Wood River, and Camas Creek drainages in central Idaho totaling approximately 8,300 square kilometers (see Figure 1). The primary focus of the study is on the water resources of the Big Wood River basin, which includes the Big Wood River and Camas Creek (approx. 6,000 square kilometers). However, approximately 100 square kilometers in the Little Wood River drainage are irrigated from the Big Wood River so that basin is included in the study area for the purpose of simulating irrigation although otherwise it is not studied in detail. The study area lies within portions of Blaine, Camas, Elmore, Gooding and Lincoln counties and the major population centers include Ketchum, Sun Valley, Hailey, Bellevue, Fairfield, and Gooding.

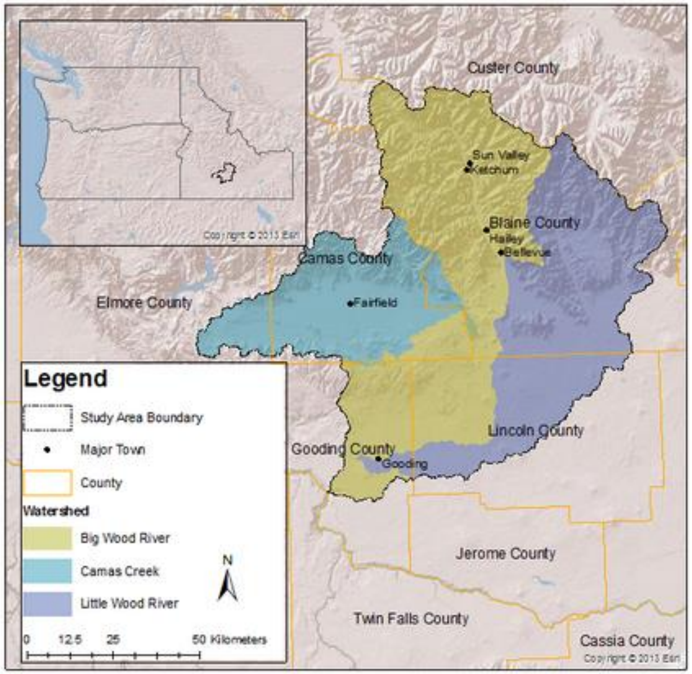


Figure 1. Map of Study Area.

Land ownership in the study area is approximately 66% public and 34% private. The majority (58%) of the public land is managed by the U.S. Bureau of Land Management (BLM) with the U.S. Forest Service overseeing 35% and the State of Idaho managing approximately 5% (U.S. Bureau of Land Management, Idaho State Office, Geographic Sciences, 2009).

Questions, comments, observations, recommendations